Simulation Exercise - Statistical Inference Assignment

Nathaniel Landi

Sunday, September 21, 2014

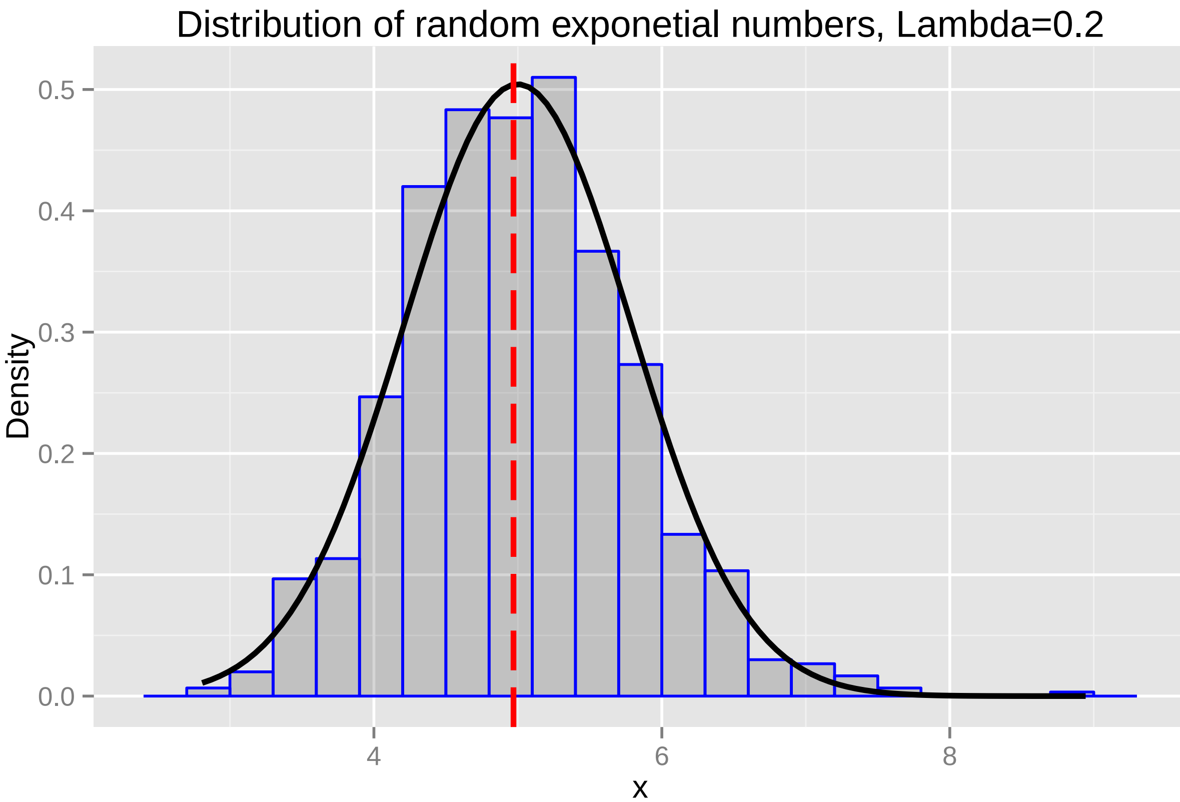
The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. In this assignment lambda = 0.2 was set for all of the simulations and the distribution of averages of 40 exponential(0.2)s was investigated. For conclusive result, a thousand or so simulated averages of 40 exponentials was performed

```
library(ggplot2)

lambda <- 0.2
n_sim <- 1000
samplesize_n <- 40
set.seed(820)

simdata <- data.frame(replicate(n_sim, mean(rexp(samplesize_n,lambda))))
names(simdata) <- c("x")
head(simdata)
```

```
##       x
## 1 5.750
## 2 3.808
## 3 4.058
## 4 3.999
## 5 4.313
## 6 4.418
```

The distribution of sample means is as follows.

```
library(ggplot2)
distribution <- ggplot(simdata, aes(x = x)) +
                geom_histogram(alpha = .20,
                               binwidth=.3,
                               colour = "blue",
                               aes(y = ..density..)) +
  stat_function(fun=dnorm,
                size = 1,
                args=list(
                  mean=mean(simdata$x),
                  sd=sd(simdata$x)))+
                  labs(title="Distribution of random exponetial numbers, Lambda=0.2",
                  y="Density") +
                  geom_vline(aes(xintercept=median(x, na.rm=T)),
                             color="red",
                             linetype="longdash",
                             size=1)
print(distribution)
```



The statistics needed to evaluate the distribution were calculated:

Mean:

```
simmean <- mean(simdata$x)
simmean
```

```
## [1] 4.999
```

Standard Deviation:

```
simsd <- sd(simdata$x)
simsd
```

```
## [1] 0.7909
```

Variance:

```
simvar <- var(simdata$x)
simvar
```

```
## [1] 0.6256
```

Expected Mean:

```
expmean = (1/lambda)
expmean
```

```
## [1] 5
```

Expected Standard Deviation:

```
expsd = (1/lambda)/sqrt(samplesize_n)
expsd
```

```
## [1] 0.7906
```

Expected Variance:

```
expvar = ((1/lambda)/sqrt(samplesize_n))^2
expvar
```

```
## [1] 0.625
```

A) The actual simulated mean is very close to the expected mean.

B) Variability of the distibution: Similarly the simulated standard deviation and variances are close to their expected values.

In accordance with the central limit theorem, the sample averages follow a normal distribution and can be seen from the above histogram and theoretical line.

The figure above shows that the density computed using the histogram and the normal density plotted with theoretical mean and variance values. This plot shows that the distribution is approximately normal.

Next we evaluate the coverage of the confidence interval for $1/\lambda = \bar{X} \pm 1.96 \frac{S}{\sqrt{n}}$.

```
Confidenceinterval <- round((simmean +( c(-1, 1) * 1.96 * simsd / sqrt(length(simdata$x))) ),2)
Confidenceinterval
```

```
## [1] 4.95 5.05
```

We can see that the average of the samples falls within this interval at least 95% of the time.

The simulation can be run with higher iterations 10000 instead of 1000 to see how the simulated statistic vary from the theoretical expected statistic as a training exercise.